Message

From: McCord, James [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=MCCORD, JAMES]
Sent: 4/25/2019 11:37:16 AM
To: Sullivan, Kate [Sullivan.Kate@epa.gov]
Subject: RE: NH MM5 Data Matrix Draft
Attachments: Chain of custody PFAS Run 1 Samples to LCMS Lab 3 21 2019.xls

I recorded the sample IDs and Descriptions from the label, but they are matched on the Chain of Custody form that Dennis worked up for me after the transfer. I found it in my emails, is attached. The two second extracts I recorded as SampleID_2

--
James McCord

From: Sullivan, Kate
Sent: Wednesday, April 24, 2019 2:51 PM
To: McCord, James <mccord.james@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

Thanks James
I'll get this started and will send for you to contribute to the draft.

Did you ever take that snapshot of your notebook with the crosswalk to the sample ids?

kate

From: McCord, James
Sent: Monday, April 22, 2019 8:46 AM
To: Sullivan, Kate <Sullivan.Kate@epa.gov>
Cc: Strynar, Mark <Strynar.Mark@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

The .xlsx file is probably better to work from. Meant to attach that instead.
--
James McCord

From: McCord, James
Sent: Monday, April 22, 2019 8:43 AM
To: Sullivan, Kate <Sullivan.Kate@epa.gov>
Cc: Strynar, Mark <strynar.mark@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

Kate,

Updated the first MM5 set with my manual annotations, a confidence flag, and a column for "priority" based on my understanding of how we have been reporting feature interest. These can be sorted and color coded to match.

Do you know the concordance of the filenames to the samples? I believe Tabor put new numbers on them and I have the descriptions but not the direct correspondence.

Description of the flag columns below:

Name

Best predicted chemical identity based on all diagnostic evidence. Loosely we label in order: Manual identification > MS/MS database match (mzCloud) > Exact mass database match (EPA Comptox Dashboard) > Formula prediction only (no Name)

Priority

1 = confirmed PFAS chemical based on manual examination of fragmentation spectrum and/or computerized match against a reference spectrum

2 = likely PFAS based on formula generation, manually examined for feasibility

3 = probably PFAS-like compound on the basis of negative mass defect in > 0.85, < 0.05 range

4 = unlikely PFAS on the basis of available MS information OR duplicate PFAS ion generated as instrumental artefact (fragments, dimers, and/or adducts when parent can be observed)

Confidence (based on Schymanski scale (DOI: 10.1021/es5002105)

1 = Confirmed by comparison with reference chemical

2a = Likely structure based on computerized spectrum match

2b = Likely structure based on manual interpretation of MS/MS spectrum

3 = Tentative candidate or MS data insufficient for unequivocal identification beyond class (i.e. PFAS chemical)

4 = Formula level identification only (MS/MS unavailable)

5 = Exact mass level identification only (no predicted formula)

Formula

Assigned formula based on highest certainty assignment (Database Match > Predicted Formula)

Score

Database match score

F_flag

TRUE = MS/MS contains diagnostic fragments for a perfluorinated substructure

rt_flag

TRUE = RT falls into a poorly resolved elution region (i.e. very early or very late) integrations and spectra are suspect

Lowabun_flag

TRUE = Passed computerized abundance filter, but spectra are of poor quality due to weak signal, assignments are suspect

md_flag

TRUE = Mass Defect is "PFAS-like"

From: Sullivan, Kate
Sent: Friday, April 12, 2019 3:30 PM
To: McCord, James <mccord.james@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

Sure. A confidence column is a good thing to call it. I guess you can always come back by finding the mass and retention time.

Kate

From: McCord, James
Sent: Friday, April 12, 2019 1:59 PM
To: Sullivan, Kate <Sullivan.Kate@epa.gov>
Cc: Strynar, Mark <Strynar.Mark@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

I haven't explicitly applied the ratings yet. For the last paper we published I did what you suggested and had a "Confidence" column with a Rating. I can provide that in the final version, along with the explanation.

Over time I would like to improve the database I am using so that when we find something novel once I can recognize it in later samples, so ideally our confidences will improve over time. But that's a non-trivial process.
--
James McCord

From: Sullivan, Kate
Sent: Friday, April 12, 2019 12:35 PM
To: McCord, James <mccord.james@epa.gov>
Cc: Strynar, Mark <Strynar.Mark@epa.gov>
Subject: RE: NH MM5 Data Matrix Draft

Nice James. I like what you're doing here.

However, I don't see where the applied Schmansky rating is. When we're all done mechanically and manually sorting and evaluating, shouldn't there be a column variable called something like "CLASSIFICATION" that each row has some assigned number ranging from 1 to 5?

Kate

From: McCord, James
Sent: Friday, April 12, 2019 8:19 AM
To: Sullivan, Kate <Sullivan.Kate@epa.gov>
Cc: Strynar, Mark <Strynar.Mark@epa.gov>
Subject: NH MM5 Data Matrix Draft

Kate,

Per our skype call I am forwarding you a draft version of the output from the MM5 NTA analysis. Columns A-K are feature metadata, remaining columns are integrated area counts for the extracted ion chromatograms of the features. Columns are named by the Sample ID

The Name column is based on a tiered reporting approach based on which matches I trust the most, this is indicated in the Match Type column:
1) Manual confirmation of name/structure (not yet performed)
2) Name based on MS/MS fragmentation match against the mzCloud database, match score reporting in the Score column (minimum allowed score = 50% match)
3) Mass/Formula match against DSSTox (no MS/MS level matching)
4) No name (formula only)
5) No name, no formula prediction

These IDs loosely match the Schymanski classifications I mentioned in the call. I attached the associated paper.

I also provided a few flags for filtering if we choose to use them later, values of TRUE mean the feature is probably not a high priority, descriptions are:

1) rt_flag TRUE = compound elutes early/late and ID is questionable due to coeluting species, normally not worth following up on
2) lowabun_flag TRUE = compound max abundance is below 50000, difficult to acquire good MS data on and hard to follow up on confirmation/normally not a major product
3) PEG_like TRUE = Part of a homologous series with organic repeating units (CH2O,CH2CH2O,etc.), very common contaminants and ubiquitous chemical series that are not typically of high tox risk, and are probably not fluorinated (this is experimental, and the first time I have used it on a dataset)
4) notPFASlike_flag TRUE = mass defect falls outside the PFAS range (> 0.85, < 0.05), unlikely to be a PFAS based on mass defect alone

My goal is to confirm and annotate the assignments starting from the top by maxArea, and also include a Notes column for my commentary on new things and artefacts. I'll stop at some reasonable level, but some of the abundant features are in-source artefacts, like the 6:1 fluorotelomer acid on line 4, and I would like to note that.

Let me know if this format works for you, if you need more explanation, or any changes you suggest.

--
James McCord